Image Sequence Analysis

Marcel Worrying
worring@science.uva.nl

Intelligent Sensory Information Systems

University of Amsterdam
Introduction: sample applications

- Information and Communication Technology
  - archives of documentaries, film, or training material, video conferencing
- Surveillance
  - guarding of gates, parking lot management
- Autonomous vehicles
  - automatically driving a car on the road, Mars explorations
- Visual inspection
  - crop monitoring, car crash analysis
- Medical diagnostics
  - heart wall motion, cell growth
Categorization

- There is a wide range of different applications, hence we have to categorize them into classes of applications.
- Categorization dimensions
  - acquisition device
  - object class
  - motion class
  - application class
Image sequence acquisition

- Acquired with a camera
  - camera mounted on a robot, train, car etc.
  - professional film equipment
  - (digital) consumer cameras
- Digitized from video tape
  - archives
  - acquired with a camera, but stored on videotape
- Acquired with special medical equipment
  - MRI, echo
- Camera mounted on a microscope
  - time-lapse microscopy
Motion types: some dimensions

- What is moving?
  - only the objects
  - only the camera
  - both

- In which space is the movement?
  - 2D, or 3D-space

- Is there a projection involved?
  - Are the individual frames in the image sequence a 2D-projection of a 3D-scene or is a full 2D or 3D space available?
Object types and their associated motion

- Rigid objects
  - objects with fixed shapes that do not change when they are moving
  - examples: cars, industrial products
Object types II

- Non-rigid objects
  - objects which can change their shape in time
  - examples: biological cells, organs, rubber objects
Object types III

- Articulated objects
  - objects composed of a set of connected rigid sub-objects with limited freedom for changing shape
  - examples: humans, robots
Object types IV

- Fluid objects
  - objects which do not have a clearly defined boundary
  - examples: moving clouds, explosions, wind in the trees
Applications of increasing complexity

- Detection of motion presence
  - detect whether somewhere in the image there is an object moving
- Locating motion
  - detect where in the image this object is moving
- Measurement of velocity
  - measure in which direction and with what speed the object is moving
- Tracking
  - follow the object through the motion
- Interpretation of motion
  - give a semantic interpretation of the movement
Basic analysis methods

- Rigid objects
  - solution method: optic flow and motion segmentation
- Non-rigid objects
  - solution method: optic flow in conjunction with snakes/deformable models
- Articulated objects
  - solution method: optic flow in conjunction with specialized objects models
- Fluids
  - solution method: optic flow based methods

*Focus: generic issues in optic flow, with an emphasis on analyzing rigid objects observed with a camera*
Object motion

Object

Pinhole camera (fixed position)

Object (moved to new position)

Pinhole camera

Movement in the image plane (projection of true motion)
Optic flow: definitions

- Optic flow
  - apparent motion of an object in the video
- Optic flow field
  - a 2D-vector field where at each position in the image an optic flow vector has been computed
- 3D-motion vector
  - movement of an object in 3D space
- Image flow
  - 2D projection of 3D-motion vector
Optic flow: examples

- Camera effects (ideal case)
  - dollying/crane left
  - zoom in
  - pan right
  - tracking
  - fast craning in arbitrary direction
  - tilt up
  - no movement
Optic flow: observations

General observation

- optic flow is often not the same as image flow
- examples of differences:
  - rotating sphere with constant light source yielding image flow and a corresponding optic flow, versus a static sphere with a rotating light source which has no image flow, but yields the same optic flow as above
  - camera movement and object movement can be indistinguishable
Bottom-up versus top-down analysis

- **Bottom-up**
  - estimation of local motion properties in the time sequence
  - integration of the measures into complete flowfields for the individual objects, or for the whole image

- **Top-down**
  - first analyze the main motion of the whole frame i.e. apparent motion of the static background using parameterized motion model, in effect finding camera motion and/or zooms
  - from there find objects which can have their own motion model as parts in the sequence not conforming to the above found motion model
Local estimation: aperture problem

- Inherent problem of local motion estimation
  - only normal motion i.e. motion perpendicular to the edge can be estimated, at corners no problem
- Each method for local optic flow estimation has to deal in some way with the aperture problem

true image flow
measured flow
due to local estimation
General local estimation scheme

- Feature measurement
  - extraction of basic measurements as gradients, spatio-temporal derivatives, local correlation surfaces etc.

- Feature integration
  - integration of the above measures to produce a complete 2D-flow field
    - involves assumptions about the smoothness of the underlying flowfield, attempting to solve the aperture problem, often uses a confidence measure to focus propagation on the most reliable measures (i.e. propagating from the corners)

- Selection
  - based on the confidence criterion
    - only relevant measurements are retained for further processing
Correlation based (example Anandan)

Based on the following equation, with $\Omega$ some neighborhood:

$$v(x) = \arg \min_v \sum_{w \in \Omega} \left( I_1(x + w) + I_2(x + w + v) \right)^2$$

Try several motion vectors and select the one for which difference in intensity values in the second image is minimal, can also be applied to derived images like Laplace.

yields estimation for motion vector at each position
Optic flow computation

- The (sum of squared differences) surface
  - the 2D-surface that you get when you plot the difference you found at each position
Confidence measures

- Derived from the shape of SSD surface around minimum
  - flat: low confidence (uniform region)
  - valley: medium confidence (edge)
  - peaked: high confidence (corner)
Feature measurement

- assumption pure translation of image with conservation of intensity $I$
- with velocity vector $\mathbf{v}$ at pixel position $\mathbf{x}$ this yields
  - $I(\mathbf{x},t) = I(\mathbf{x} - \mathbf{v}t,0)$
- taking the derivative with respect to $t$ on both sides yields the gradient constraint equation
  - $\nabla I(\mathbf{x},t) \cdot \mathbf{v} + I_t(\mathbf{x},t) = 0$
Feature integration

- the gradient constraint equation yields one equation in the two unknown elements of the velocity vector $v$ (due to the aperture problem)
- solution is to gather the estimations in a larger neighborhood $\Omega$
- assume constant velocity within $\Omega$ and solve the following equation in least squares sense, with $W$ a weighting factor gradually decreasing from the center we have:

$$
\sum_{x \in \Omega} W(x)^2 (\nabla I(x,t).v + I_t(x,t))^2
$$
Gradient based

- Feature selection
  - based on the eigenvectors of $A^T W^2 A$, where $A$ contains the intensity gradients at every position in the neighborhood, indicating whether:
    - gradients align: hence we are at an edge and can hence compute normal flow only
    - gradients do not align: hence we are at a corner and true optic flow can be computed
Multi-scale analysis and selection

- Feature integration
  - assume smoothness of entire flowfield
  - make local assumption that velocity vectors change very slowly
  - propagate reliable information from coarse to more detailed levels

- Selection
  - based directly on the confidence factors
Optic flow estimation result
Global motion estimation

- Image alignment
  - find correspondence between two subsequent frames in the time sequence
  - requires to find the relative position of the two frames (in 3D world space)
    - with respect to each other and with respect to the camera

- Assumptions
  - camera motion is the dominant motion
    - no scene activity apart from camera motion
    - activity small part of the image field
  - camera motion can be estimated from some dominant plane in the 3D-world
    - in general the background
Global motion estimation

Object
Plane in the world

Pinhole camera (fixed position)

Parameterized transformation

Movement in the image plane

Object
Plane in the world

Pinhole camera (changed position)
Global motion model

- Any 2D motion field of a plane moving in 3D space can be written in the form
  - \( u(x) = p_1 x + p_2 y + p_5 + p_7 x^2 + p_8 xy \)
  - \( v(x) = p_3 x + p_4 y + p_6 + p_7 xy + p_8 y^2 \)
  - where \( x = (x,y) \) and \( u,v \) are the components of the motion vector \( \mathbf{v} \)

- Characteristics
  - describes the change of position of each point in the image which is part of the planar surface
  - each instantiation of the 8 parameters corresponds to a different flow field
Global motion model

- Examples with origin at the center of gravity of the object

\[ u(x) = 1, v(x) = 2 \quad \text{i.e.} \]
\[ p5 \text{ and } p6 \text{ set all other } 0 \]

\[ u(x) = 0, v(x) = 0. \quad 1y \quad \text{i.e. only } p4 \text{ non } 0 \]
Image alignment

- Align the two image frames based on the motion model
  - corresponds to a virtual camera movement
  - if the two images are properly aligned i.e. p1-8 are correct then there should be no motion in the common part

$\begin{align*}
I_i & \quad \text{parameterized transform p1-8} \\
I_{i+1} & \quad \text{Frames in common coordinate system}
\end{align*}$
Goal: find the best values of p1-8

- Choose initial transform
- Try to maximize matching function by choosing different p1-8
- Use a spatial pyramid: first coarse alignment than more precise
Application: spatio-temporal segmentation

- Decomposition
  - segmentation of the images sequence into objects using the observation that parts of an objects move in a coherent way

- Indexing
  - we will look at methods that parameterize the motion of an object so they form a proper basis for further processing

- Methods
  - top-down or bottom-up
Spatio-temporal segmentation: Top-down

1. Compute global motion model
   - Put pixels that fit the model well (small residuals) into one layer
2. Collect points with high residuals
Spatio-temporal segmentation: bottom-up

Combine pixels with the same color

Merge groups of pixels if they can be described using the same motion model
Residuals

- Causes of residuals
  - change of content
    - object movement
    - illumination change
  - inaccuracies
    - misalignments
    - interpolation errors during warping
    - noise
Result of segmentation
Conclusion

- Optic flow
  - optic flow forms the basis for almost all of the applications in image sequence analysis
- From application to method selection
  - there is a lot of variety in image sequence analysis problems and applications, categorization of your problem according to the criteria is an important step in selection of an appropriate method
References

- B.K.P. Horn: Computer Vision, MIT Press, 1986